(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 905 008 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
03.11.2021 Bulletin 2021/44

(21) Application number: **21180094.1**

(22) Date of filing: **13.12.2017**

(51) Int Cl.:
G06F 3/01 *(2006.01)* ***G06F 3/03*** *(2006.01)*
G06K 9/00 *(2006.01)* ***G06K 9/46*** *(2006.01)*

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
17934888.3 / 3 610 358

(71) Applicant: **SZ DJI Technology Co., Ltd.**
Shenzhen, Guangdong 518057 (CN)

(72) Inventors:
- **ZHOU, You**
  Shenzhen (CN)
- **LIU, Jie**
  Shenzhen (CN)
- **ZHU, Zhenyu**
  Shenzhen (CN)

(74) Representative: **Appelt, Christian W.**
Boehmert & Boehmert
Anwaltspartnerschaft mbB
Pettenkoferstrasse 22
80336 München (DE)

Remarks:
This application was filed on 17-06-2021 as a divisional application to the application mentioned under INID code 62.

(54) DEPTH INFORMATION BASED GESTURE DETERMINATION FOR MOBILE PLATFORMS

(57) Determining the pose or gesture of one or more subjects in an environment adjacent to a mobile platform, and associated systems and methods are disclosed herein. A representative method includes identifying candidate regions from depth data representing the environment based on depth connectivity criterion, associating disconnected regions for identifying multiple pose components of the subject, determining a spatial relationship between the identified pose components, and cause generation of a controlling command based thereon.

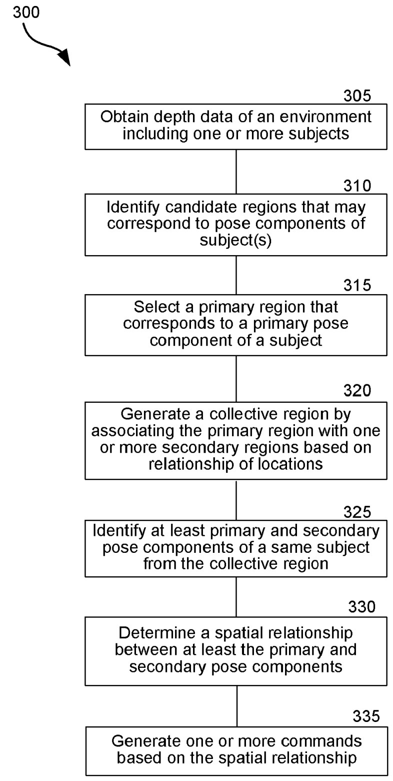
300
305
Obtain depth data of an environment including one or more subjects
310
Identify candidate regions that may correspond to pose components of subject(s)
315
Select a primary region that corresponds to a primary pose component of a subject
320
Generate a collective region by associating the primary region with one or more secondary regions based on relationship of locations
325
Identify at least primary and secondary pose components of a same subject from the collective region
330
Determine a spatial relationship between at least the primary and secondary pose components
335
Generate one or more commands based on the spatial relationship

FIG. 3

EP 3 905 008 A1

Printed by Jouve, 75001 PARIS (FR)

Description

TECHNICAL FIELD

[0001] The present technology is generally directed to determining the pose or gesture of one or more subjects, such as one or more human users in a three-dimensional (3D) environment adjacent to a mobile platform.

BACKGROUND

[0002] The environment surrounding a mobile platform can typically be scanned or otherwise detected using one or more sensors. For example, the mobile platform can be equipped with a stereo vision system (e.g., a "stereo camera") to sense its surrounding environment. A stereo camera is typically a type of camera with two or more lenses each having a separate image sensor or film frame. When taking photos/videos with the two or more lenses at the same time but from different angles, the difference between the corresponding photos/videos provides a basis for calculating depth information (e.g., distance from objects in the scene to the stereo camera). As another example, the mobile platform can be equipped with one or more LiDAR sensors, which typically transmit a pulsed signal (e.g. laser signal) outwards, detect the pulsed signal reflections, and determine depth information about the environment to facilitate object detection and/or recognition. However, inaccuracies exist in different depth-sensing technologies, which can affect various higher level applications such as pose and/or gesture determination.

SUMMARY

[0003] The following summary is provided for the convenience of the reader and identifies several representative embodiments of the disclosed technology.

[0004] In some embodiments, a computer-implemented method for determining a pose of a subject within an environment includes identifying a plurality of candidate regions from base depth data representing the environment based, at least in part, on at least one depth connectivity criterion, determining a first region from a subset of the candidate regions as corresponding to a first pose component of the subject, and selecting one or more second regions from the subset of candidate regions to associate with the first region based, at least in part, on relative locations of the first region and the one or more second regions. The method also includes identifying the first pose component and at least one second pose component of the subject from a collective region including the first region and the one or more associated second regions, determining a spatial relationship between the identified first pose component and the identified at least one second pose component, and causing generation of a controlling command for execution by a mobile platform based, at least in part, on the determined spatial relationship.

[0005] In some embodiments, the base depth data is generated based, at least in part, on images captured by at least one stereo camera. In some embodiments, the base depth data includes a depth map calculated based, at least in part, on a disparity map and intrinsic parameters of the at least one stereo camera. In some embodiments, the method further comprises determining a depth range where the subject is likely to appear in the base depth data. In some embodiments, the plurality of candidate regions are identified within the range of depth. In some embodiments, the base depth data includes at least one of unknown, invalid, or inaccurate depth information.

[0006] In some embodiments, the at least one depth connectivity criterion includes at least one of a depth threshold or a change-of-depth threshold. In some embodiments, two or more of candidate regions are disconnected from one another due, at least in part, to unknown, invalid, or inaccurate depth information in the base depth data. In some embodiments, the method further comprises selecting the subset of the candidate regions based, at least in part, on first baseline information regarding at least one pose component of the subject. In some embodiments, the first baseline information indicates an estimated size of the at least one pose component. In some embodiments, the first baseline information is based, at least in part, on prior depth data.

[0007] In some embodiments, determining the first region is based, at least in part, on second baseline information regarding the first pose component of the subject. In some embodiments, the second baseline information indicates an estimated location of the first pose component. In some embodiments, the second baseline information is based, at least in part, on prior depth data. In some embodiments, selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on non-depth information regarding the environment. In some embodiments, the non-depth information includes two-dimensional image data that corresponds to the base depth data.

[0008] In some embodiments, selecting the one or more second regions to associate with the first region comprises distinguishing candidate regions potentially corresponding to the subject from candidate regions potentially corresponding to at least one other subject. In some embodiments, selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on estimated locations of a plurality of joints of the subject. In some embodiments, selecting one or more second regions from the subset of candidate regions to associate with the first region at least partially discounts an effect of unknown, invalid, or inaccurate depth information in the base depth data.

[0009] In some embodiments, the subject is a human being. In some embodiments, the first pose component and the at least one second pose component are body parts of the subject. In some embodiments, the first pose component is a torso of the subject and the at least one second pose component is a hand of the subject. In some embodiments, identifying the first pose component and at least one second pose component of the subject comprises detecting a portion of the collective region based, at least in part, on a measurement of depth. In some embodiments, detecting a portion of the collective region comprises detecting a portion closest in depth to the mobile platform. In some embodiments, identifying the at least one second pose component is based, at least in part, on a location of the detected portion relative to a grid system.

[0010] In some embodiments, identifying the at least one second pose component comprises identifying at least two second pose components. In some embodiments, identifying the first pose component comprises removing the identified at least one second pose component from the collective region. In some embodiments, identifying the first pose component is based, at least in part, on a threshold on width.

[0011] In some embodiments, determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more geometric attributes of the first and/or second pose components. In some embodiments, the one or more geometric attributes include a location of centroid. In some embodiments, determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining comprises determining one or more vectors pointing between portions of the first pose component and the at least one second pose component.

[0012] In some embodiments, the mobile platform includes at least one of an unmanned aerial vehicle (UAV), a manned aircraft, an autonomous car, a self-balancing vehicle, a robot, a smart wearable device, a virtual reality (VR) head-mounted display, or an augmented reality (AR) head-mounted display. In some embodiments, the method further comprises controlling a mobility function of the mobile platform based, at least in part, on the controlling command.

[0013] Any of the foregoing methods can be implemented via a non-transitory computer-readable medium storing computer-executable instructions that, when executed, cause one or more processors associated with a mobile platform to perform corresponding actions, or via a vehicle including a programmed controller that at least partially controls one or more motions of the vehicle and that includes one or more processors configured to perform corresponding actions.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014]

Figure 1A is an illustrative 2D image of an environment in front of a mobile platform and Figure 1B is an illustrative depth map corresponding to the 2D image of Figure 1A, in accordance with some embodiments of the presently disclosed technology.

Figure 2A is an illustrative 2D image of a subject (e.g., human user) and Figure 2B is a corresponding depth map with defects, in accordance with some embodiments of the presently disclosed technology.

Figure 3 is a flowchart illustrating a method for determining a pose or gesture of a subject, in accordance with some embodiments of the presently disclosed technology.

Figure 4A illustrates a 2D image of multiple subjects (e.g., two human dancers) and Figure 4B illustrates subject features (e.g. joints) determined from the 2D image, as well as the grouping and mapping of the subject features to individual subjects, in accordance with some embodiments of the presently disclosed technology.

Figures 5A-5D illustrate a process for identifying a secondary pose component (e.g., hand) of a subject, in accordance with some embodiments of the presently disclosed technology.

Figures 6A and 6B illustrate a process for identifying a primary pose component (e.g., torso) of the subject, in accordance with some embodiments of the presently disclosed technology.

Figure 7 illustrates examples of mobile platforms configured in accordance with various embodiments of the presently disclosed technology.

Figure 8 is a block diagram illustrating an example of the architecture for a computer system or other control device that can be utilized to implement various portions of the presently disclosed technology.

DETAILED DESCRIPTION

1. Overview

[0015] Poses or gestures are natural communication modes for human users, which provide a promising direction for a human-computer interface applicable to various mobile platforms. However, it can be technologically difficult and computationally intensive to determine the pose or gesture of subject(s) based on two-dimensional (2D) image information, and any resulting distance information may not be accurate. Depth information (e.g., provided by stereo camera, LiDAR, and/or other sensors) provides measurements in another dimension, which can be used to improve the efficiency and/or efficacy of pose determination. However, inaccuracies or other imperfections in the depth information may occur for various reasons. For example, a stereo camera may not be able to provide depth information for a texture-less subject, some part(s) of the environment may be overexposed or underexposed, and certain shapes, orientations or colocations of objects may cause imperfect depth detection.

[0016] The presently disclosed technology uses a region-based method to discount or eliminate the negative effect of imperfections in depth information, identify pose components of one or more subjects, and determine spatial relationships among the pose components for generating various controlling commands based thereon. An illustrative method can include analyzing depth information (e.g., including identifying depth inaccuracies or other imperfections) and determining candidate regions which potentially includes at least a portion of a pose component (e.g., torso, hand, or the like). The candidate regions can be determined based on various criteria of depth-discontinuity, some of which may be caused by the depth inaccuracies or other imperfections. The illustrative method can include generating a collective region by grouping together a subset of the candidate regions based on their spatial or logical relationship, thereby "reconnecting" certain candidate regions that were separated due to the depth inaccuracies or imperfections. The collective region can be further analyzed for identifying one or more pose components therein.

[0017] Several details describing structures and/or processes that are well-known and often associated with mobile platforms (e.g., UAVs or other types of movable platforms) and corresponding systems and subsystems, but that may unnecessarily obscure some significant aspects of the presently disclosed technology, are not set forth in the following description for purposes of clarity. Moreover, although the following disclosure sets forth several embodiments of different aspects of the presently disclosed technology, several other embodiments can have different configurations or different components than those described herein. Accordingly, the presently disclosed technology may have other embodiments with additional elements and/or without several of the elements described below with reference to Figures 1A-8.

[0018] Figures 1A-8 are provided to illustrate representative embodiments of the presently disclosed technology. Unless provided for otherwise, the drawings are not intended to limit the scope of the claims in the present application.

[0019] Many embodiments of the technology described below may take the form of computer- or controller-executable instructions, including routines executed by a programmable computer or controller. The programmable computer or controller may or may not reside on a corresponding scanning platform. For example, the programmable computer or controller can be an onboard computer of the mobile platform, or a separate but dedicated computer associated with the mobile platform, or part of a network or cloud based computing service. Those skilled in the relevant art will appreciate that the technology can be practiced on computer or controller systems other than those shown and described below. The technology can be embodied in a special-purpose computer or data processor that is specifically programmed, configured or constructed to perform one or more of the computer-executable instructions described below. Accordingly, the terms "computer" and "controller" as generally used herein refer to any data processor and can include Internet appliances and handheld devices (including palm-top computers, wearable computers, cellular or mobile phones, multi-processor systems, processor-based or programmable consumer electronics, network computers, mini computers and the like). Information handled by these computers and controllers can be presented at any suitable display medium, including an LCD (liquid crystal display). Instructions for performing computer- or controller-executable tasks can be stored in or on any suitable computer-readable medium, including hardware, firmware or a combination of hardware and firmware. Instructions can be contained in any suitable memory device, including, for example, a flash drive, USB (universal serial bus) device, and/or other suitable medium. In particular embodiments, the instructions are accordingly non-transitory.

2. Representative Embodiments

[0020] As discussed above, a stereo camera or other sensor can provide data for obtaining depth information (e.g., measurements of distance between different portions of a scene and the sensor) for an environment that surrounds or is otherwise adjacent to, but does not necessarily abut, a mobile platform. For example, Figure 1A is an illustrative 2D image of an environment in front of a mobile platform and Figure 1B is an illustrative depth map corresponding to the 2D image of Figure 1A, in accordance with some embodiments of the presently disclosed technology. As shown in Figure 1A, the environment includes a subject (e.g., a human user), various other objects, a floor, and walls. The depth map of Figure 1B illustratively uses grayscale to represent the depth (e.g., distance to the mobile platform or the observing sensor) of different portions of the environment. For example, the darker a pixel appears in the depth map of Figure 1B, the deeper (i.e., farther away) the portion of the environment corresponding to the pixel is located. In various embodiments, a depth map can be represented as a color graph, point cloud, or other forms.

[0021] Various imperfections, such as inaccuracies, defects, and/or unknown or invalid values, can exist in the depth information about an environment. For example, Figure 2A is an illustrative 2D image of a subject (e.g., human user) and Figure 2B is a corresponding depth map with defects, in accordance with some embodiments of the presently disclosed technology. As can be seen, the subject in Figure 2A is extending an arm to express a pose/gesture. A portion 202 of the extended arm lacks texture (e.g., appears mostly in white, without variance). In this case, the stereo camera generated depth map of Figure 2B cannot accurately reflect the depth of the arm portion 202. Illustratively, in Figure 2B, a region 204 that corresponds to the arm portion 202 includes unknown or invalid depth data, thereby incorrectly separating a torso portion 206 from a hand portion 208, both of which in fact belong to a same subject. A person of skill in the art can appreciate that various other forms of imperfections may exist in sensor-generated depth information about an environment.

[0022] Figure 3 is a flowchart illustrating a method 300 for determining a pose or gesture of a subject, despite imperfections, in accordance with some embodiments of the presently disclosed technology. The method can be implemented by a controller (e.g., an onboard computer of a mobile platform, an associated computing device, and/or an associated computing service). The method 300 can use various combinations of depth data, non-depth data, and/or baseline information at different stages to achieve pose/gesture determinations. As discussed above in paragraph [0023], the method 300 method can include analyzing depth information, determining candidate regions, as well as generating a collective region for identifying one or more pose components therein.

[0023] At block 305, the method includes obtaining depth data of an environment (e.g., adjacent to the mobile platform) including one or more subjects (e.g., human users). The depth data can be generated based on images captured by one or more stereo cameras, point clouds provided by one or more LiDAR sensors, and/or data produced by other sensors carried by the mobile platform. In the embodiments in which stereo camera(s) are used, the depth data can be represented as a series of time-sequenced depth maps that are calculated based on corresponding disparity maps and intrinsic parameters of the stereo camera(s). In some embodiments, median filters or other preprocessing operations can be applied to source data (e.g., disparity maps) before they are used as a basis for generating the depth data, in order to reduce noise or otherwise improve data quality.

[0024] In some embodiments, the method includes determining a depth range in which one or more subjects are likely to appear in the depth data corresponding to a particular time (e.g., a most recently generated frame of a depth map). Illustratively, the depth range can be determined based on baseline information derived from prior depth data (e.g., one or more frames of depth maps generated immediately prior to the most recent frame). The baseline information can include a depth location (e.g., 2 meters in front of the mobile platform) of a subject that was determined (e.g., using method 300 in a prior iteration) based on the prior depth data. The depth range (e.g., a range between 1.5 meters and 2.5 meters) can be set to cover the determined depth location with certain margins. In some embodiments, the baseline information can be defined independent of the existence of prior depth data. For example, an initial iteration of the method 300 can use a predefined depth range (e.g., a range between 1 meter and 3.5 meters in front of the mobile platform). The controller can generate a reference depth map (e.g., Figure 2B) using only the depth information within the depth range. In some embodiments, the depth range determination is not performed and the reference depth map can include all or a subset of depth information corresponding to the particular time.

[0025] At block 310, the method includes identifying candidate regions that may correspond to pose components (e.g., the user's or other subject's torso, hand, or the like). Illustratively, the controller analyzes the reference depth map to identify all the regions that can be separated from one another based on one or more depth connectivity criteria (e.g., a depth threshold or a change-of-depth threshold). For example, each identified candidate region includes pixels that are connected to one another on the depth map. The pixels in each region do not vary in depth beyond a certain threshold. Each candidate region is disconnected from any other candidate regions in depth, and in some embodiments, also disconnected in one or two other dimensions. As discussed previously, in some cases, the disconnection is caused by unknown, invalid, or inaccurate depth information.

[0026] The controller can select a subset of the candidate regions based on baseline information regarding at least one pose component of the subject. The baseline information can indicate an estimated size of pose component(s) (e.g., size of a hand), based on prior depth data. Illustratively, the controller can estimate a quantity n_{vaild} of pixels on the reference depth map that correspond to an average size of a hand, using (1) the likely depth location (e.g., 2 meters in front of the mobile platform) of the subject as estimated at block 305 and (2) a previously determined relationship between a pixel quantity (e.g., 100 pixels) and an average hand size at a known depth location (e.g., 1 meter in front of the mobile platform). In this example, the estimated pixel quantity n_{vaild} can be 25 pixels based on operation of geometric transformations. In some embodiments, a margin (e.g., 20%) or other weighting factor can be applied to reduce the risk of false negatives, thus the estimated pixel quantity n_{vaild} can be further reduced, for example, to 20 pixels. Accordingly, potentially more candidate regions can be selected and included in the subset.

[0027] The estimated size of the pose component(s) can be used to filter the candidate regions to generate a more relevant subset for identifying pose component(s) in a more efficient manner. Illustratively, the controller compares the size (e.g., number of pixels) of each candidate region with a valid size range (e.g., between n_{vaild} and a multiple of n_{vaild}) defined based on the estimated size, and filters out candidate regions having sizes that fall outside of the valid size range. In some embodiments, the valid size range can be determined based on estimated sizes of two or more pose components.

[0028] At block 315, the method includes selecting from the subset of candidate regions a primary region that potentially corresponds to a primary pose component (e.g., torso) of a subject. In some embodiments, the controller determines the primary region based on baseline information (e.g., location and/or size) regarding the primary pose component of the subject. Similarly, the baseline information used here can be derived from prior depth data (e.g., prior reference depth map(s) where the primary pose component was determined using method 300 in prior iteration(s)). For example, if the subject's torso was detected from a prior frame of depth information, the controller can determine a centroid point of the previously determined torso portion and map the centroid point to the current reference depth map. The controller can label a candidate region as the primary region if the centroid point lands inside the candidate region. Alternatively, the controller can select a primary region based on region size (e.g., selecting the largest candidate region within the subset as corresponding to the torso of the subject). It should be noted that the primary region may include a portion of the primary pose component, multiple pose components, non-pose components (e.g., head of a human user in some cases) of the subject, and/or objects other than the subject.

[0029] At block 320, the method includes generating a collective region by associating the primary region with one or more secondary regions of the subset based on their relative locations. In some embodiments, this can be achieved based on the analysis of non-depth information regarding the environment. As discussed previously, candidate regions can be disconnected from one another due to unknown, invalid, or inaccurate depth information. Non-depth information (e.g., corresponding 2D image(s) of the environment) can be used to re-connect or otherwise associate disconnected regions. Illustratively, joints or other applicable subject features can be determined from 2D images. Based on the spatial and/or logical relationship among the determined subject features, they can be grouped in an ordered fashion and/or mapped to corresponding subject(s). The grouping and/or mapping can be based on baseline information regarding the subject features (e.g., the wrist joint and shoulder joint must be connected via an elbow joint). In some embodiments, the controller can implement a Part Affinity Fields (PAFs)-based method to achieve the feature grouping and/or mapping. For example, Figure 4A illustrates a 2D image of multiple subjects (e.g., two human dancers), and Figure 4B illustrates subject features (e.g. joints) determined from the 2D image, as well as the grouping and mapping of subject features to individual subjects, in accordance with some embodiments of the presently disclosed technology. As illustrated in Figure 4B, the grouping and mapping of subject features can distinguish the two subjects from each other, and thus avoid incorrectly mapping feature(s) of a first subject to a second subject. The controller can then project the subject features to the reference depth map and associate the primary region (e.g., that includes one or more features of a subject) with one or more secondary regions (e.g., each including one or more features of the same subject) in accordance with the subject feature grouping and/or mapping.

[0030] Because analyzing non-depth information may consume a considerable amount of computational resources, in some embodiments, the controller generates the collective region based on the distance between and/or the relative sizes of the primary region and other candidate regions. For example, the controller can select a threshold number of secondary regions that (1) have a size within the valid size range and (2) are located close enough to the primary region to be associated with the primary region. Once the collective region, including the primary and secondary regions, is constructed, other candidate regions can be filtered out or otherwise excluded from further processing.

[0031] At block 325, the method includes identifying at least a primary pose component (e.g., torso) and a secondary pose component (e.g., hand) of a same subject from the collective region. Illustratively, the controller analyzes a distribution of depth measurements using baseline information regarding the pose component(s).

[0032] For example, Figures 5A-5D illustrate a process for identifying a secondary pose component (e.g., hand) of a subject (e.g., a human user), in accordance with some embodiments of the presently disclosed technology. For purposes of illustration, Figures 5A-5D do not show depth information (e.g., in grayscale). The baseline information regarding pose component(s) can indicate that a hand of a subject is likely to be a pose component that has a shortest distance to the mobile platform. Therefore, as illustrated in Figure 5A, the controller can search for and identify a closest point 502 within the collective region 510. With reference to Figure 5B, using the closest point as a seed point, the controller can identify a portion 504 of the collective region, for example, by applying a flood-fill method that is initiated at the closest point 502. Various other applicable methods can be employed to identify the portion 504 based on the closest point 502, and various depth thresholds or change-of-depth thresholds can be used therein. The identified portion 504 can include a hand of the subject and potentially other body part(s) (e.g., arm).

[0033] In some embodiments, the controller can divide or partition the reference depth map, in order to determine the secondary pose component in a more efficient and/or accurate manner. For example, a grid system can be used to further determine the hand of the subject within the identified portion 504. As illustrated in Figure 5C, an example grid system 520 can divide the reference depth map into 9 grid blocks 531-539. The gridlines of the grid system 520 can be defined based on baseline information regarding the primary pose component (e.g., torso) or other pose components of the subject, and may themselves be non-uniform. For example, the upper horizontal gridline 522 can be located at a height of a torso centroid estimated from one or more previous frames of depth data; the lower horizontal gridline 524 can be located at a margin distance from the upper horizontal gridline 522; and the left and right vertical gridlines 526, 528 can be defined based on a torso width (with certain margin value) estimated from one or more previous frames of depth data.

[0034] The gridlines form various grid blocks that can be processed individually (in sequence or in parallel) based on the grid block's location, size, or other attributes. Illustratively, the controller can scan the identified portion 504 in a direction or manner based on a location of the grid block the portion 504 falls into. For example, if the identified portion 504 falls into any of grid blocks 531-536 or 538, it is more likely that the hand of the subject is pointing upwards. Therefore, at least for purposes of computational efficiency, the controller can scan the identified portion 504, pixel row by pixel row, in an up-down direction. As another example, if the identified portion falls into either grid block 537 or 539, it is more likely that the hand of the subject is pointing left or pointing right. Therefore, the controller can scan the identified portion, pixel column by pixel column, in a left-right or right-left direction, respectively. As illustrated in Figure 5D, an estimated boundary 542 of the hand can be located where an increase in depth between two or more adjacent pixel rows (or columns) exceeds a threshold. The set of scanned pixels 540 within the estimated hand boundary 542 can be identified as corresponding to a hand of the subject.

[0035] In some embodiments, the hand identification is further corroborated or revised based on 2D image data that corresponds to the reference depth map. For example, the hand boundaries can be revised based on texture and/or contrast analysis of the 2D image data. In some embodiments, the controller removes the identified first hand from the collective region and repeats the process of Figures 5A-5D to identify the second hand of the subject.

[0036] Figures 6A and 6B illustrate a process for identifying a primary pose component (e.g., torso) of the subject, in accordance with some embodiments of the presently disclosed technology. The collective region may include multiple pose components of the subject, non-pose components (e.g., head or legs in certain cases) of the subject, and/or objects other than the subject. Accordingly, the controller further processes the collective region to particularly identify the primary pose component. Illustratively, the controller can remove the identified secondary pose component(s) (e.g., hand(s)) and analyze the rest of the collective region 620. For example, the controller can efficiently identify the torso of the subject based on widths at various heights of the remaining collective region 620. As illustrated in Figure 6A, the remaining collective region 620 can be examined line by line at different heights. The controller can compare widths of the collective region at different heights to a torso width range (e.g., between 2 and 6 times the width of the hand) and identify the subject's torso based thereon. With reference to Figure 6A, an example line 602 indicates a width 604 that is outside the torso width range (e.g., smaller than twice of the hand width), while another example line 606 indicates a width 608 that falls within the torso width range. Accordingly, as illustrated in Figure 6B, the controller identifies a larger rectangular area 610 as the torso of the subject. The area 610 excludes the upper half of the subject's head (where line 602 resides).

[0037] At block 330, the method includes determining a spatial relationship between at least the identified primary pose component and secondary pose component(s). The controller can determine one or more geometric attributes (e.g., the centroid, contour, shape, or the like) of the primary and/or secondary pose components. The controller can determine one or more vectors pointing between portions of the primary pose component and the secondary pose component(s). At block 335, the method includes generating one or more commands based on the determined spatial relationship. Illustratively, the determined geometric attributes and/or the vectors can be used to define a direction, speed, acceleration, and/or rotation, which serves as a basis for generating command(s) for controlling the mobile platform's next move(s). For example, the controller can generate a command that causes the mobile platform to stop moving when a user's hand is raised. As another example, the controller can generate a command that causes the mobile platform to move toward a point in space that resides on a line defined by the centroids of the identified torso and hand of the subject. In various embodiments, the method 300 can be implemented in response to obtaining each frame of depth data, in response to certain events (e.g., the mobile platform detecting presence of one or more subjects), and/or based on user commands.

[0038] Figure 7 illustrates examples of mobile platforms configured in accordance with various embodiments of the presently disclosed technology. As illustrated, a representative mobile platform as disclosed herein may include at least one of an unmanned aerial vehicle (UAV) 702, a manned aircraft 704, an autonomous car 706, a self-balancing vehicle 708, a terrestrial robot 710, a smart wearable device 712, a virtual reality (VR) head-mounted display 714, or an augmented reality (AR) head-mounted display 716.

[0039] Figure 8 is a block diagram illustrating an example of the architecture for a computer system or other control device 800 that can be utilized to implement various portions of the presently disclosed technology. In Figure 8, the computer system 800 includes one or more processors 805 and memory 810 connected via an interconnect 825. The interconnect 825 may represent any one or more separate physical buses, point to point connections, or both, connected by appropriate bridges, adapters, or controllers. The interconnect 825, therefore, may include, for example, a system bus, a Peripheral Component Interconnect (PCI) bus, a HyperTransport or industry standard architecture (ISA) bus, a small computer system interface (SCSI) bus, a universal serial bus (USB), IIC (I2C) bus, or an Institute of Electrical and Electronics Engineers (IEEE) standard 674 bus, sometimes referred to as "Firewire."

[0040] The processor(s) 805 may include central processing units (CPUs) to control the overall operation of, for example, the host computer. In certain embodiments, the processor(s) 805 accomplish this by executing software or firmware stored in memory 810. The processor(s) 805 may be, or may include, one or more programmable general-purpose or special-purpose microprocessors, digital signal processors (DSPs), programmable controllers, application specific integrated circuits (ASICs), programmable logic devices (PLDs), or the like, or a combination of such devices.

[0041] The memory 810 can be or include the main memory of the computer system. The memory 810 represents any suitable form of random access memory (RAM), read-only memory (ROM), flash memory, or the like, or a combination of such devices. In use, the memory 810 may contain, among other things, a set of machine instructions which, when executed by processor 805, causes the processor 805 to perform operations to implement embodiments of the presently disclosed technology.

[0042] Also connected to the processor(s) 805 through the interconnect 825 is a (optional) network adapter 815. The network adapter 815 provides the computer system 800 with the ability to communicate with remote devices, such as the storage clients, and/or other storage servers, and may be, for example, an Ethernet adapter or Fiber Channel adapter.

[0043] The techniques described herein can be implemented by, for example, programmable circuitry (e.g., one or more microprocessors) programmed with software and/or firmware, or entirely in special-purpose hardwired circuitry, or in a combination of such forms. Special-purpose hardwired circuitry may be in the form of, for example, one or more application-specific integrated circuits (ASICs), programmable logic devices (PLDs), field-programmable gate arrays (FPGAs), etc.

[0044] Software or firmware for use in implementing the techniques introduced here may be stored on a machine-readable storage medium and may be executed by one or more general-purpose or special-purpose programmable microprocessors. A "machine-readable storage medium," as the term is used herein, includes any mechanism that can store information in a form accessible by a machine (a machine may be, for example, a computer, network device, cellular phone, personal digital assistant (PDA), manufacturing tool, any device with one or more processors, etc.). For example, a machine-accessible storage medium includes recordable/non-recordable media (e.g., read-only memory (ROM); random access memory (RAM); magnetic disk storage media; optical storage media; flash memory devices; etc.), etc.

[0045] The term "logic," as used herein, can include, for example, programmable circuitry programmed with specific software and/or firmware, special-purpose hardwired circuitry, or a combination thereof.

[0046] Some embodiments of the disclosure have other aspects, elements, features, and/or steps in addition to or in place of what is described above. These potential additions and replacements are described throughout the rest of the specification. Reference in this specification to "various embodiments," "certain embodiments," or "some embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the disclosure. These embodiments, even alternative embodiments (e.g., referenced as "other embodiments") are not mutually exclusive of other embodiments. Moreover, various features are described which may be exhibited by some embodiments and not by others. Similarly, various requirements are described which may be requirements for some embodiments but not other embodiments. For example, some embodiments uses depth information generated from stereo camera(s), while other embodiments can use depth information generated from LiDAR(s), 3D-ToF, or RGB-D. Still further embodiments can use depth information generated from a combination of sensors.

[0047] To the extent any materials incorporated by reference herein conflict with the present disclosure, the present disclosure controls. Various embodiments have been described. The present invention may also (alternatively) be described by the below numbered aspects 1 to 100.

ASPECTS

[0048]

1. A computer-implemented method for determining a gesture of a user within a three-dimensional (3D) environment based, at least in part, on depth information, the method comprising:

   determining a depth range where the user is likely to appear in a current depth map of the environment based, at least in part, on one or more prior depth maps;

   generating a reference depth map based, at least on part, on filtering the current depth map using the range of depth;

   identifying a plurality of candidate regions from the reference depth map, wherein changes in depth within each of the plurality of candidate regions do not exceed a threshold value and wherein the plurality of candidate regions are disconnected from one another;

   selecting a subset from the plurality of candidate regions, wherein each selected candidate region has a size within a threshold range and wherein the threshold range is based on an estimated size of a hand of the user relative to the target depth map;

   determining a primary region from the subset

based, at least in part, on a location and/or size corresponding to a torso of the user;
associating the primary region with one or more target regions of the subset as potentially corresponding to body parts of the user based, at least in part, on relative locations of the primary region and the one or more target regions;
identifying a hand of the user from a collective region that includes the primary region and the one or more target regions;
identifying the torso of the user after filtering out the identified hand from the collective region;
determining one or more vectors representing a spatial relationship between at least the identified torso and identified hand of the user; and
controlling a movement of a mobile platform based, at least in part, on the determined one or more vectors.

2. A computer-implemented method for determining a pose of a subject within an environment based, at least in part, on depth information, the method comprising:

identifying a plurality of candidate regions from base depth data representing the environment based, at least in part, on at least one depth connectivity criterion;
determining a first region from a subset of the candidate regions based, at least in part, on an estimation regarding a first pose component of the subject;
selecting one or more second regions from the subset of candidate regions to associate with the first region based, at least in part, on relative locations of the first region and the one or more second regions;
identifying the first pose component and at least one second pose component of the subject from a collective region including the first region and the one or more associated second regions;
determining a spatial relationship between the identified first pose component and the identified at least one second pose component; and
causing generation of a controlling command for execution by a mobile platform based, at least in part, on the determined spatial relationship.

3. The method of aspect 2, wherein the base depth data is generated based, at least in part, on images captured by at least one stereo camera.

4. The method of aspect 3, wherein the base depth data includes a depth map calculated based, at least in part, on a disparity map and intrinsic parameters of the at least one stereo camera.

5. The method of any of aspects 2-4, further comprising determining a depth range where the subject is likely to appear in the base depth data.

6. The method of aspect 5, wherein the plurality of candidate regions are identified within the range of depth.

7. The method of any of aspects 2-6, wherein the base depth data includes at least one of unknown, invalid, or inaccurate depth information.

8. The method of aspect 2, wherein the at least one depth connectivity criterion includes at least one of a depth threshold or a change-of-depth threshold.

9. The method of any of aspects 2 or 8, wherein two or more of candidate regions are disconnected from one another due, at least in part, to unknown, invalid, or inaccurate depth information in the base depth data.

10. The method of any of aspects 2, 8, or 9, further comprising selecting the subset of the candidate regions based, at least in part, on first baseline information regarding at least one pose component of the subject.

11. The method of aspect 10, wherein the first baseline information indicates an estimated size of the at least one pose component.

12. The method of aspect 10, wherein the first baseline information is based, at least in part, on prior depth data.

13. The method of aspect 2, wherein determining the first region is based, at least in part, on second baseline information regarding the first pose component of the subject.

14. The method of aspect 13, wherein the second baseline information indicates an estimated location of the first pose component.

15. The method of aspect 13, wherein the second baseline information is based, at least in part, on prior depth data.

16. The method of aspects 2, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on non-depth information regarding the environment.

17. The method of aspect 16, wherein the non-depth information includes two-dimensional image data that corresponds to the base depth data.

18. The method of any of aspects 2, 16, or 17, wherein selecting the one or more second regions to associate with the first region comprises distinguishing candidate regions potentially corresponding to the subject from candidate regions potentially corresponding to at least one other subject.

19. The method of any of aspects 2 or 16-18, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on estimated locations of a plurality of joints of the subject.

20. The method of any of aspects 2 or 16-19, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region at least partially discounts an effect of unknown, invalid, or inaccurate depth information in the base depth data.

21. The method of aspect 2, wherein the subject is a human being.

22. The method of any of aspects 2 or 21, wherein the first pose component and the at least one second pose component are body parts of the subject.

23. The method of any of aspects 2, 21, or 22, wherein the first pose component is a torso of the subject and the at least one second pose component is a hand of the subject.

24. The method of any of aspects 2 or 21-23, wherein identifying the first pose component and at least one second pose component of the subject comprises detecting a portion of the collective region based, at least in part, on a measurement of depth.

25. The method of aspect 24, wherein detecting a portion of the collective region comprises detecting a portion closest in depth to the mobile platform.

26. The method of any of aspects 24 or 25, wherein identifying the at least one second pose component is based, at least in part, on a location of the detected portion relative to a grid system.

27. The method of any of aspects 2 or 21-26, wherein identifying the at least one second pose component comprises identifying two second pose components.

28. The method of any of aspects 2 or 21-27, wherein identifying the first pose component comprises removing the identified at least one second pose component from the collective region.

29. The method of any of aspects 2 or 21-28, wherein identifying the first pose component is based, at least in part, on a threshold on width.

30. The method of aspect 2, wherein determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more geometric attributes of the first and/or second pose components.

31. The method of aspect 30, wherein the one or more geometric attributes include a location of centroid.

32. The method of any of aspects 2, 30, or 31, wherein determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more vectors pointing between portions of the first pose component and the at least one second pose component.

33. The method of any of aspects 2-32, wherein the mobile platform includes at least one of an unmanned aerial vehicle (UAV), a manned aircraft, an autonomous car, a self-balancing vehicle, a robot, a smart wearable device, a virtual reality (VR) head-mounted display, or an augmented reality (AR) head-mounted display.

34. The method of any of aspects 2-33, further comprising controlling a mobility function of the mobile platform based, at least in part, on the controlling command.

35. A non-transitory computer-readable medium storing computer-executable instructions that, when executed, cause one or more processors associated with a mobile platform to perform actions, the actions comprising:

identifying a plurality of candidate regions from base depth data representing the environment based, at least in part, on at least one depth connectivity criterion;
determining a first region from a subset of the candidate regions based, at least in part, on an estimation regarding a first pose component of the subject;
selecting one or more second regions from the subset of candidate regions to associate with the first region based, at least in part, on relative locations of the first region and the one or more second regions;
identifying the first pose component and at least one second pose component of the subject from a collective region including the first region and the one or more associated second regions;
determining a spatial relationship between the

identified first pose component and the identified at least one second pose component; and
causing generation of a controlling command for execution by the mobile platform based, at least in part, on the determined spatial relationship.

36. The computer-readable medium of aspect 35, wherein the base depth data is generated based, at least in part, on images captured by at least one stereo camera.

37. The computer-readable medium of aspect 36, wherein the base depth data includes a depth map calculated based, at least in part, on a disparity map and intrinsic parameters of the at least one stereo camera.

38. The computer-readable medium of any of aspects 35-37, wherein the actions further comprise determining a depth range where the subject is likely to appear in the base depth data.

39. The computer-readable medium of aspect 38, wherein the plurality of candidate regions are identified within the range of depth.

40. The computer-readable medium of any of aspects 35-39, wherein the base depth data includes at least one of unknown, invalid, or inaccurate depth information.

41. The computer-readable medium of aspect 35, wherein the at least one depth connectivity criterion includes at least one of a depth threshold or a change-of-depth threshold.

42. The computer-readable medium of any of aspects 35 or 41, wherein two or more of candidate regions are disconnected from one another due, at least in part, to unknown, invalid, or inaccurate depth information in the base depth data.

43. The computer-readable medium of any of aspects 35, 41, or 42, wherein the actions further comprise selecting the subset of the candidate regions based, at least in part, on first baseline information regarding at least one pose component of the subject.

44. The computer-readable medium of aspect 43, wherein the first baseline information indicates an estimated size of the at least one pose component.

45. The computer-readable medium of aspect 43, wherein the first baseline information is based, at least in part, on prior depth data.

46. The computer-readable medium of aspect 35, wherein determining the first region is based, at least in part, on second baseline information regarding the first pose component of the subject.

47. The computer-readable medium of aspect 46, wherein the second baseline information indicates an estimated location of the first pose component.

48. The computer-readable medium of aspect 46, wherein the second baseline information is based, at least in part, on prior depth data.

49. The computer-readable medium of aspects 35, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on non-depth information regarding the environment.

50. The computer-readable medium of aspect 49, wherein the non-depth information includes two-dimensional image data that corresponds to the base depth data.

51. The computer-readable medium of any of aspects 35, 49, or 50, wherein selecting the one or more second regions to associate with the first region comprises distinguishing candidate regions potentially corresponding to the subject from candidate regions potentially corresponding to at least one other subject.

52. The computer-readable medium of any of aspects 35 or 49-51, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on estimated locations of a plurality of joints of the subject.

53. The computer-readable medium of any of aspects 35 or 49-52, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region at least partially discounts an effect of unknown, invalid, or inaccurate depth information in the base depth data.

54. The computer-readable medium of aspect 35, wherein the subject is a human being.

55. The computer-readable medium of any of aspects 35 or 54, wherein the first pose component and the at least one second pose component are body parts of the subject.

56. The computer-readable medium of any of aspects 35, 54, or 55, wherein the first pose component is a torso of the subject and the at least one second pose component is a hand of the subject.

57. The computer-readable medium of any of aspects 35 or 54-56, wherein identifying the first pose component and at least one second pose component of the subject comprises detecting a portion of the collective region based, at least in part, on a measurement of depth.

58. The computer-readable medium of aspect 57, wherein detecting a portion of the collective region comprises detecting a portion closest in depth to the mobile platform.

59. The computer-readable medium of any of aspects 57 or 58, wherein identifying the at least one second pose component is based, at least in part, on a location of the detected portion relative to a grid system.

60. The computer-readable medium of any of aspects 35 or 54-59, wherein identifying the at least one second pose component comprises identifying at least two second pose components.

61. The computer-readable medium of any of aspects 35 or 54-60, wherein identifying the first pose component comprises removing the identified at least one second pose component from the collective region.

62. The computer-readable medium of any of aspects 35 or 54-61, wherein identifying the first pose component is based, at least in part, on a threshold on width.

63. The computer-readable medium of aspect 35, wherein determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more geometric attributes of the first and/or second pose components.

64. The computer-readable medium of aspect 63, wherein the one or more geometric attributes include a location of centroid.

65. The computer-readable medium of any of aspects 35, 63, or 64, wherein determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more vectors pointing between portions of the first pose component and the at least one second pose component.

66. The computer-readable medium of any of aspects 35-65, wherein the mobile platform includes at least one of an unmanned aerial vehicle (UAV), a manned aircraft, an autonomous car, a self-balancing vehicle, a robot, a smart wearable device, a virtual reality (VR) head-mounted display, or an augmented reality (AR) head-mounted display.

67. The computer-readable medium of any of aspects 35-66, wherein the actions further comprise controlling a mobility function of the mobile platform based, at least in part, on the controlling command.

68. A vehicle including a controller programmed to at least partially control one or more motions of the vehicle, wherein the programmed controller includes one or more processors configured to:

> identify a plurality of candidate regions from base depth data representing the environment based, at least in part, on at least one depth connectivity criterion;
> determine a first region from a subset of the candidate regions based, at least in part, on an estimation regarding a first pose component of the subject;
> select one or more second regions from the subset of candidate regions to associate with the first region based, at least in part, on relative locations of the first region and the one or more second regions;
> identify the first pose component and at least one second pose component of the subject from a collective region including the first region and the one or more associated second regions;
> determine a spatial relationship between the identified first pose component and the identified at least one second pose component; and
> cause generation of a controlling command for execution by the vehicle based, at least in part, on the determined spatial relationship.

69. The vehicle of aspect 68, wherein the base depth data is generated based, at least in part, on images captured by at least one stereo camera.

70. The vehicle of aspect 69, wherein the base depth data includes a depth map calculated based, at least in part, on a disparity map and intrinsic parameters of the at least one stereo camera.

71. The vehicle of any of aspects 68-70, wherein the one or more processors are further configured to determine a depth range where the subject is likely to appear in the base depth data.

72. The vehicle of aspect 71, wherein the plurality of candidate regions are identified within the range of depth.

73. The vehicle of any of aspects 68-72, wherein the base depth data includes at least one of unknown, invalid, or inaccurate depth information.

74. The vehicle of aspect 68, wherein the at least one depth connectivity criterion includes at least one of a depth threshold or a change-of-depth threshold.

75. The vehicle of any of aspects 68 or 74, wherein two or more of candidate regions are disconnected from one another due, at least in part, to unknown, invalid, or inaccurate depth information in the base depth data.

76. The vehicle of any of aspects 68, 74, or 75, wherein the one or more processors are further configured to select the subset of the candidate regions based, at least in part, on first baseline information regarding at least one pose component of the subject.

77. The vehicle of aspect 76, wherein the first baseline information indicates an estimated size of the at least one pose component.

78. The vehicle of aspect 76, wherein the first baseline information is based, at least in part, on prior depth data.

79. The vehicle of aspect 68, wherein determining the first region is based, at least in part, on second baseline information regarding the first pose component of the subject.

80. The vehicle of aspect 79, wherein the second baseline information indicates an estimated location of the first pose component.

81. The vehicle of aspect 79, wherein the second baseline information is based, at least in part, on prior depth data.

82. The vehicle of aspects 68, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on non-depth information regarding the environment.

83. The vehicle of aspect 82, wherein the non-depth information includes two-dimensional image data that corresponds to the base depth data.

84. The vehicle of any of aspects 68, 82, or 83, wherein selecting the one or more second regions to associate with the first region comprises distinguishing candidate regions potentially corresponding to the subject from candidate regions potentially corresponding to at least one other subject.

85. The vehicle of any of aspects 68 or 82-84, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on estimated locations of a plurality of joints of the subject.

86. The vehicle of any of aspects 68 or 82-85, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region at least partially discounts an effect of unknown, invalid, or inaccurate depth information in the base depth data.

87. The vehicle of aspect 68, wherein the subject is a human being.

88. The vehicle of any of aspects 68 or 87, wherein the first pose component and the at least one second pose component are body parts of the subject.

89. The vehicle of any of aspects 68, 87, or 88, wherein the first pose component is a torso of the subject and the at least one second pose component is a hand of the subject.

90. The vehicle of any of aspects 68 or 87-89, wherein identifying the first pose component and at least one second pose component of the subject comprises detecting a portion of the collective region based, at least in part, on a measurement of depth.

91. The vehicle of aspect 90, wherein detecting a portion of the collective region comprises detecting a portion closest in depth to the vehicle.

92. The vehicle of any of aspects 90 or 91, wherein identifying the at least one second pose component is based, at least in part, on a location of the detected portion relative to a grid system.

93. The vehicle of any of aspects 68 or 87-92, wherein identifying the at least one second pose component comprises identifying at least two second pose components.

94. The vehicle of any of aspects 68 or 87-93, wherein identifying the first pose component comprises removing the identified at least one second pose component from the collective region.

95. The vehicle of any of aspects 68 or 87-94, wherein identifying the first pose component is based, at least in part, on a threshold on width.

96. The vehicle of aspect 68, wherein determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more geometric attributes of the first and/or second pose components.

97. The vehicle of aspect 96, wherein the one or more geometric attributes include a location of centroid.

98. The vehicle of any of aspects 68, 96, or 97, wherein determining a spatial relationship between the identified first pose component and the identified at least one second pose component comprises determining one or more vectors pointing between portions of the first pose component and the at least one second pose component.

99. The vehicle of any of aspects 68-98, wherein the vehicle includes at least one of an unmanned aerial vehicle (UAV), a manned aircraft, an autonomous car, a self-balancing vehicle, a robot, a smart wearable device, a virtual reality (VR) head-mounted display, or an augmented reality (AR) head-mounted display.

100. The vehicle of any of aspects 68-99, wherein the one or more processors are further configured to control a mobility function of the vehicle based, at least in part, on the controlling command.

Claims

1. A computer-implemented method for determining a gesture of a user within a three-dimensional (3D) environment based, at least in part, on depth information, the method comprising:

determining a depth range where the user is likely to appear in a current depth map of the environment based, at least in part, on one or more prior depth maps;
generating a reference depth map based, at least on part, on filtering the current depth map using the range of depth;
identifying a plurality of candidate regions from the reference depth map, wherein changes in depth within each of the plurality of candidate regions do not exceed a threshold value and wherein the plurality of candidate regions are disconnected from one another;
selecting a subset from the plurality of candidate regions, wherein each selected candidate region has a size within a threshold range and wherein the threshold range is based on an estimated size of a hand of the user relative to the target depth map;
determining a primary region from the subset based, at least in part, on a location and/or size corresponding to a torso of the user;
associating the primary region with one or more target regions of the subset as potentially corresponding to body parts of the user based, at least in part, on relative locations of the primary region and the one or more target regions;
identifying a hand of the user from a collective region that includes the primary region and the one or more target regions;
identifying the torso of the user after filtering out the identified hand from the collective region;
determining one or more vectors representing a spatial relationship between at least the identified torso and identified hand of the user; and
controlling a movement of a mobile platform based, at least in part, on the determined one or more vectors.

2. The method of claim 1, wherein the plurality of candidate regions are identified within the range of depth.

3. The method of claim 1 or 2, wherein the depth range is determined based on one or more frames of depth maps generated immediately prior to the most recent frame.

4. The method of any of claims 1-3, wherein each candidate region is disconnected from any other candidate regions in depth.

5. The method of any of claims 1-4, wherein the primary region is determined based on baseline information, such as location and/or size, regarding the torso of the subject.

6. The method of any one of claims 1-5, further comprising:
generating the collective region based on the distance between and/or the relative sizes of the primary region and other candidate regions.

7. The method of any of claims 1 to 6, further comprising:
constructing the collective region and once the collective region is constructed, filtering out or otherwise excluding from further processing other candidate regions.

8. The method of any one of claims 1 to 7, further comprising: using the vectors to define a direction, speed, acceleration, and/or rotation, as a basis for generating a command for controlling of the movement of the mobile platform.

9. The method of any of claims 1 to 8, wherein selecting the one or more second regions to associate with the first region comprises distinguishing candidate regions potentially corresponding to the subject from candidate regions potentially corresponding to at least one other subject.

10. The method of any of claims 1 to 9, wherein selecting one or more second regions from the subset of candidate regions to associate with the first region is based, at least in part, on estimated locations of a plurality of joints of the subject.

11. The method of any one of claims 1 to 10, wherein the subject is a human being.

12. The method of any of claims 1-11, wherein the mobile platform includes at least one of an unmanned aerial vehicle (UAV), a manned aircraft, an autonomous car, a self-balancing vehicle, a robot, a smart wearable device, a virtual reality (VR) head-mounted display, or an augmented reality (AR) head-mounted display.

13. The method of claim 8, further comprising controlling a mobility function of the mobile platform based, at least in part, on the command.

14. A non-transitory computer-readable medium storing computer-executable instructions that, when executed, cause one or more processors associated with a mobile platform to perform the method of any one of the preceding claims.

15. A vehicle including a controller programmed to at least partially control one or more motions of the vehicle, wherein the programmed controller includes one or more processors configured to perform the method of any one of claims 1 to 13.

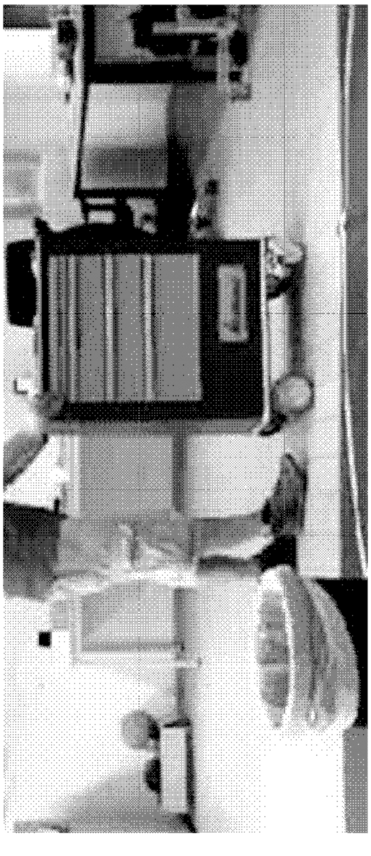

FIG. 1A

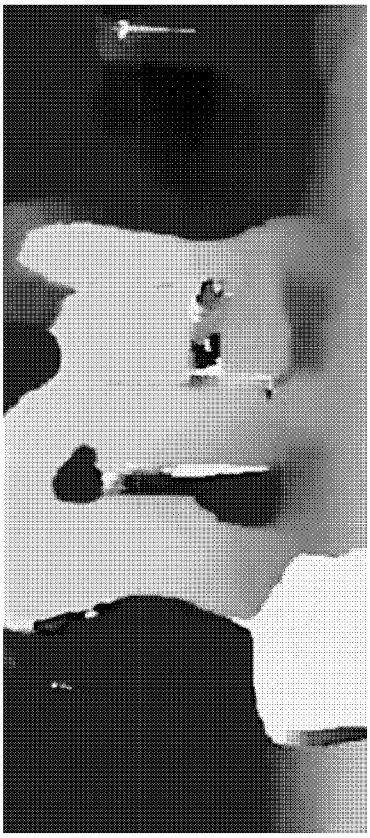

FIG. 1B

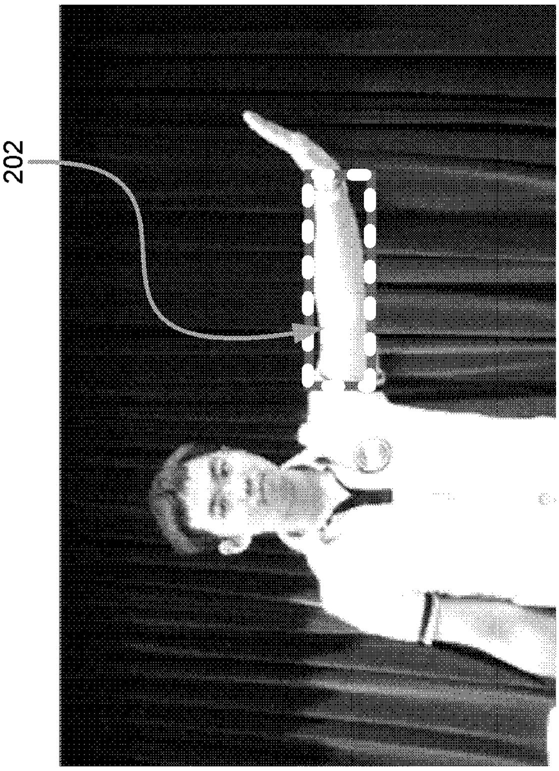

FIG. 2A

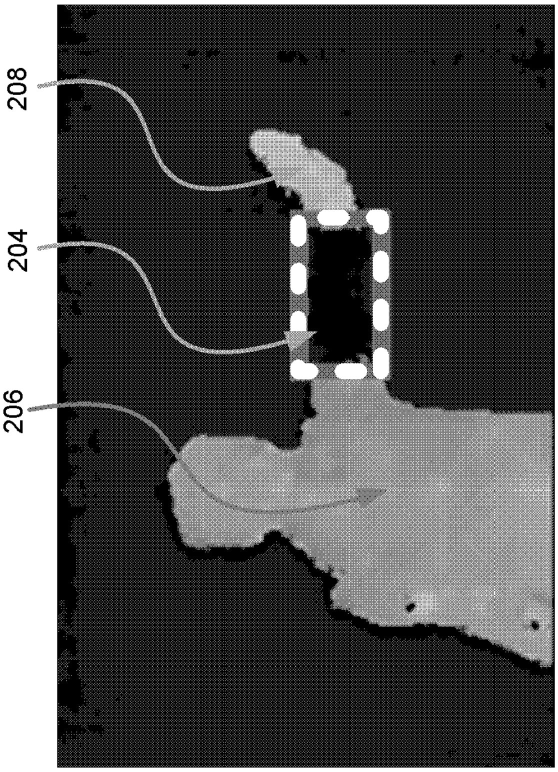

FIG. 2B

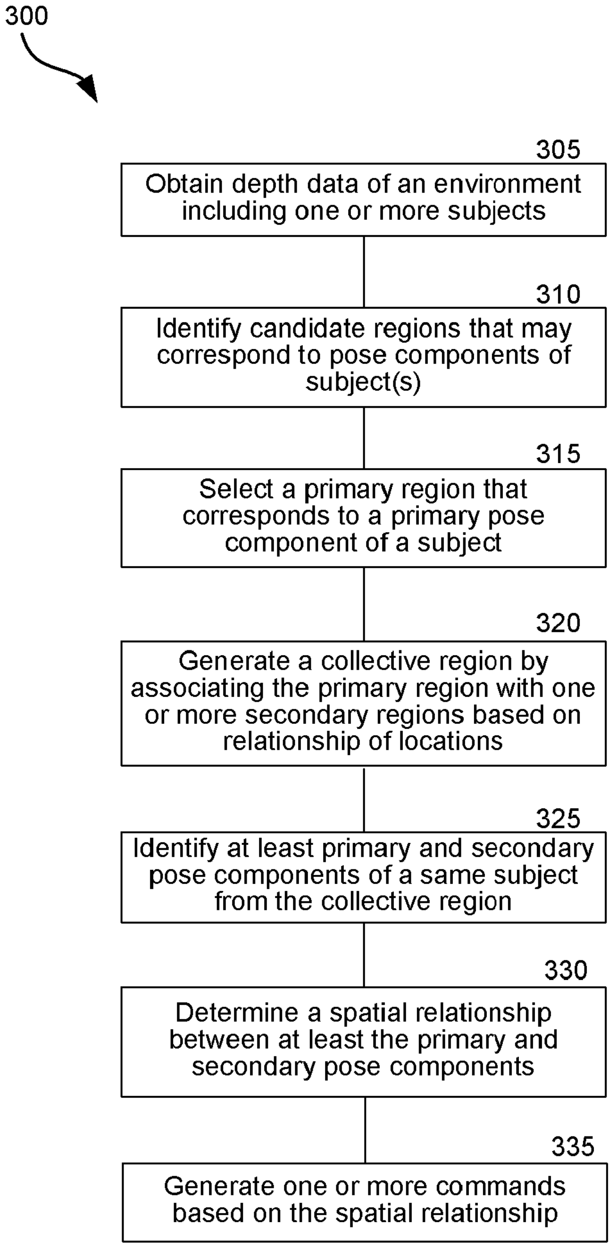
300
305
Obtain depth data of an environment including one or more subjects
310
Identify candidate regions that may correspond to pose components of subject(s)
315
Select a primary region that corresponds to a primary pose component of a subject
320
Generate a collective region by associating the primary region with one or more secondary regions based on relationship of locations
325
Identify at least primary and secondary pose components of a same subject from the collective region
330
Determine a spatial relationship between at least the primary and secondary pose components
335
Generate one or more commands based on the spatial relationship

FIG. 3

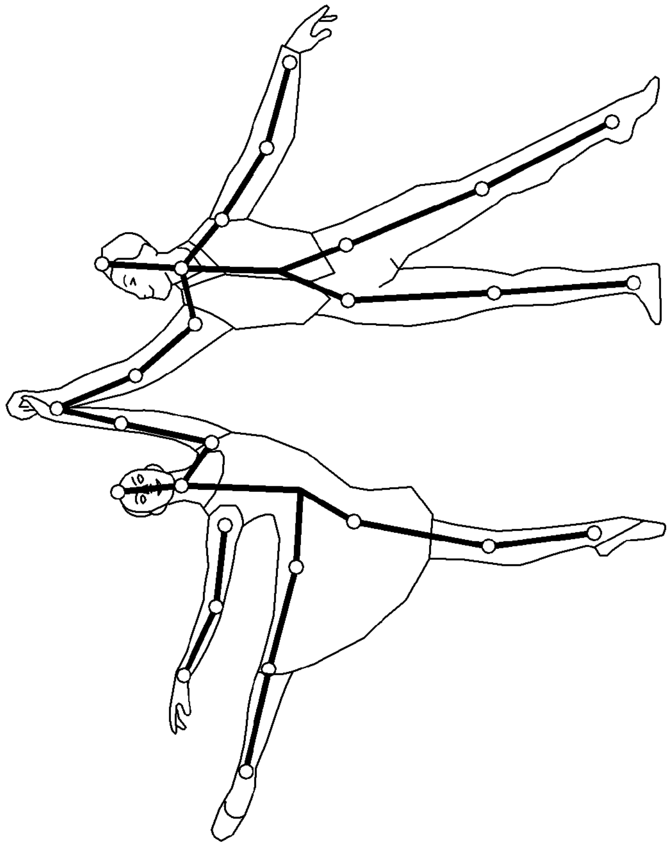

FIG. 4B

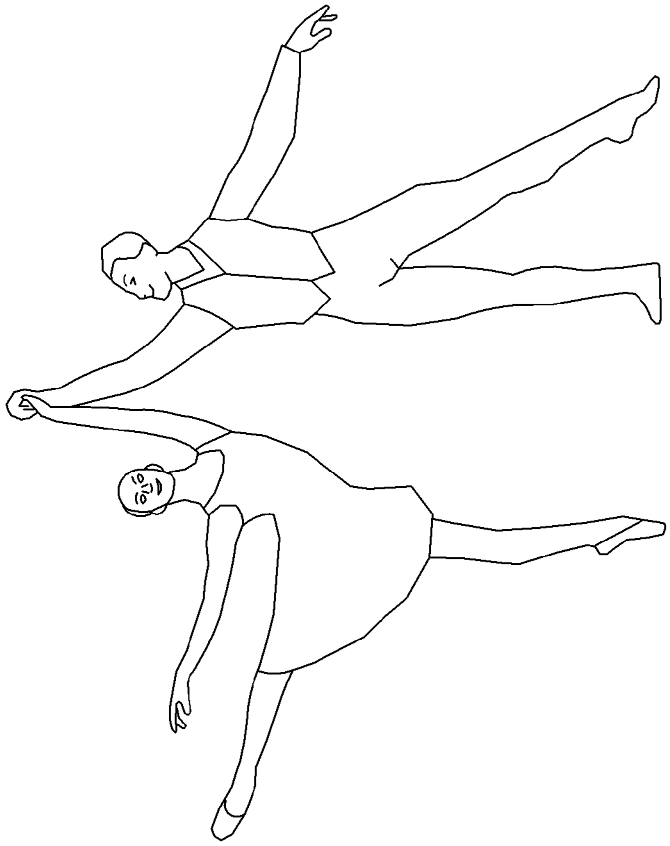

FIG. 4A

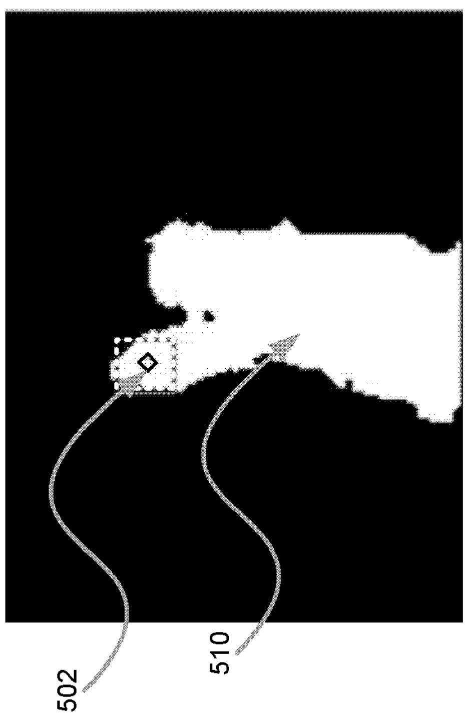

FIG. 5A

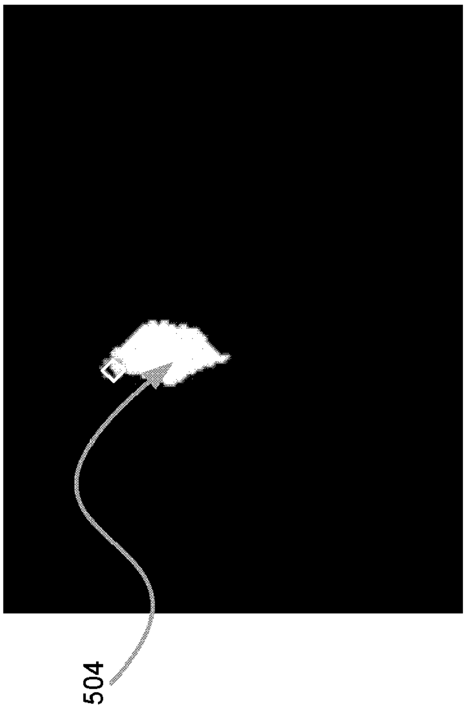

FIG. 5B

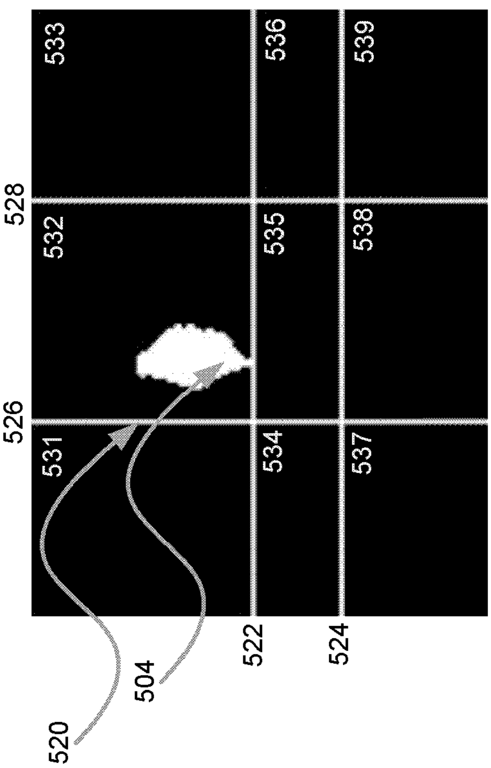

FIG. 5C

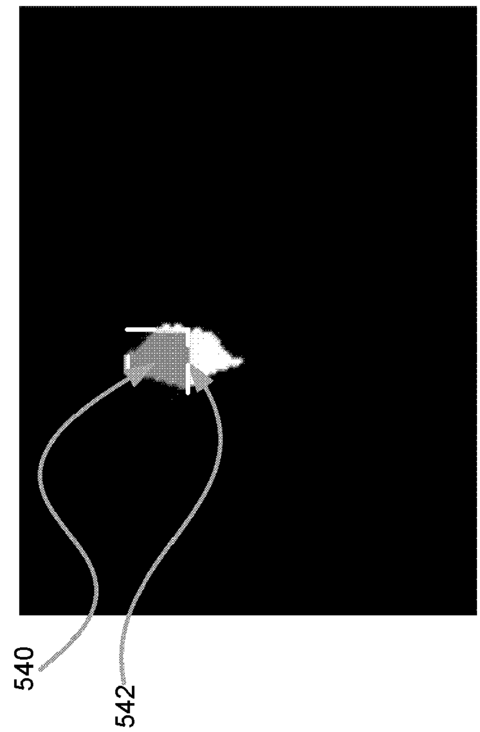

FIG. 5D

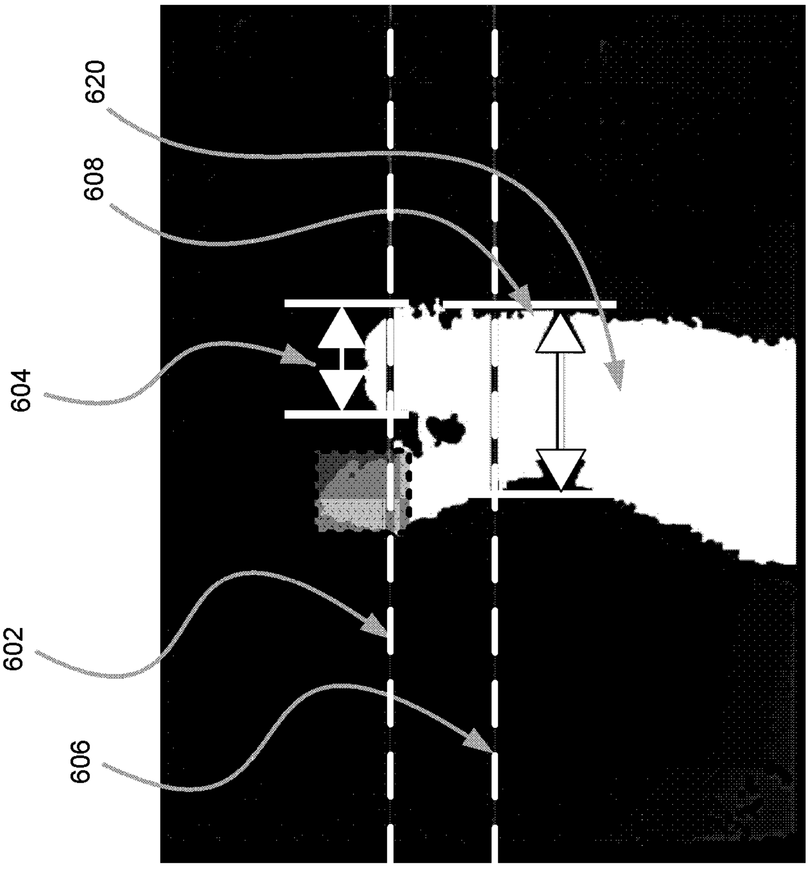

FIG. 6A

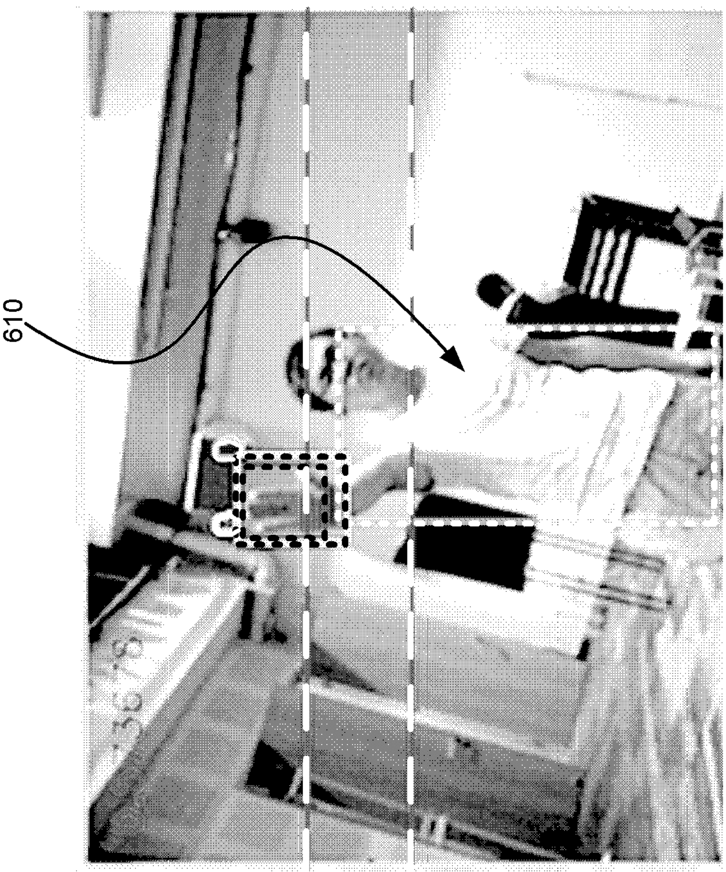

FIG. 6B

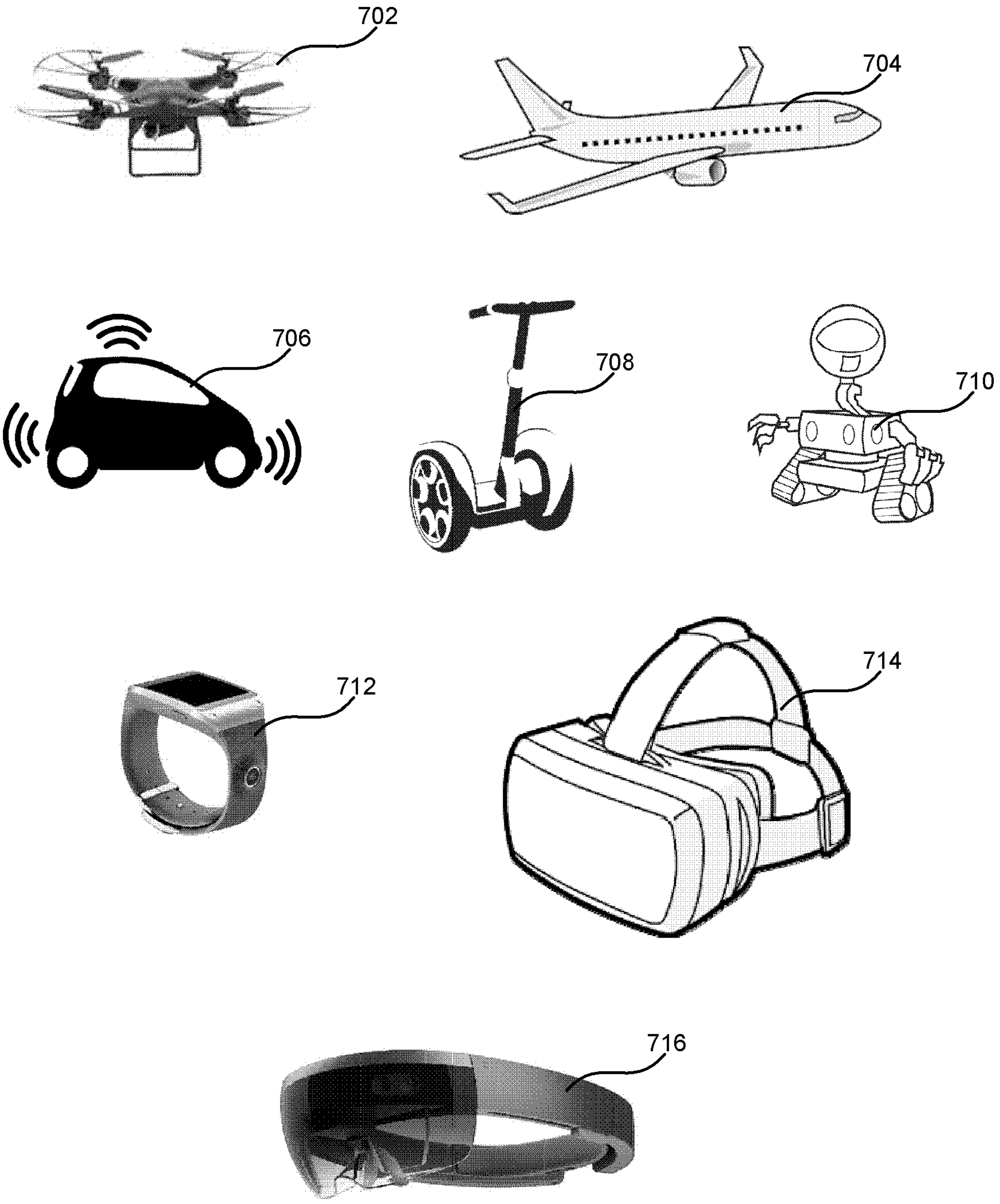

FIG. 7

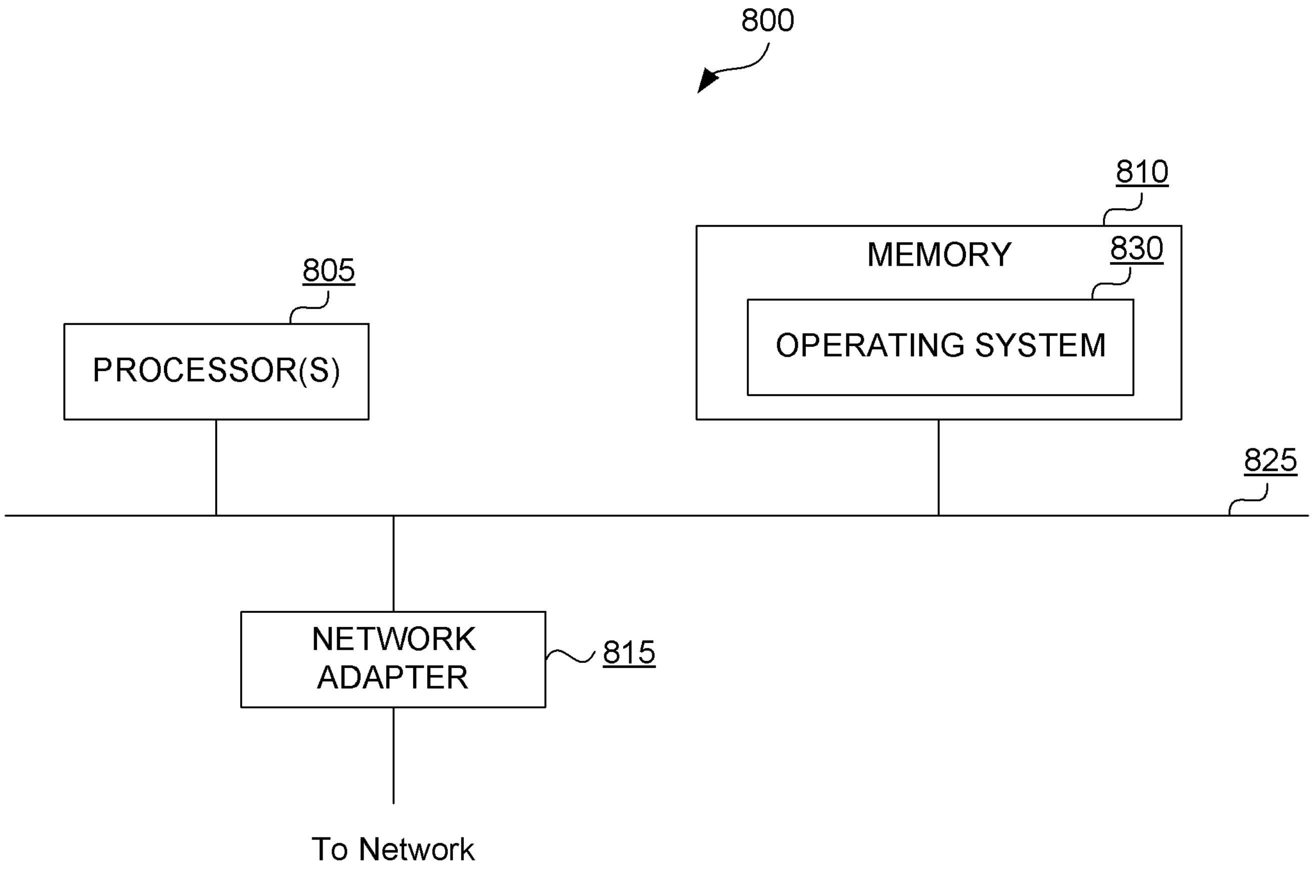
800
810
MEMORY
830
OPERATING SYSTEM
805
PROCESSOR(S)
825
NETWORK
ADAPTER
815
To Network

FIG. 8

Europäisches
Patentamt
European
Patent Office
Office européen
des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 21 18 0094

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
A	US 2014/056471 A1 (GU JIN [CA]) 27 February 2014 (2014-02-27) * abstract; figures 1-19 * * paragraphs [0002], [0014], [0015], [0039] - [0043], [0051], [0052], [0072], [0077] - [0083], [0101], [0107] - [0140], [0158] - [0162] * * paragraph [0187] * -----	1-15	INV. G06F3/01 G06F3/03 G06K9/00 G06K9/46
A	US 2016/267699 A1 (BORKE MICHAEL JAMES [US] ET AL) 15 September 2016 (2016-09-15) * abstract; figures 1-33 * * paragraphs [0102] - [0107] * -----	1,14,15	
A	US 2017/351253 A1 (YANG JIANJUN [CN] ET AL) 7 December 2017 (2017-12-07) * abstract; figures 1-4C * * paragraphs [0040], [0049], [0050], [0061], [0070] * -----	1,8, 12-15	
			TECHNICAL FIELDS SEARCHED (IPC)
			G06F G06K

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	27 September 2021	Köhn, Andreas

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
....
& : member of the same patent family, corresponding document

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 21 18 0094

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

27-09-2021

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014056471 A1	27-02-2014	US 2014056471 A1	27-02-2014
		US 2014056472 A1	27-02-2014
		WO 2014031331 A1	27-02-2014
		WO 2014031332 A1	27-02-2014
US 2016267699 A1	15-09-2016	CA 2979218 A1	15-09-2016
		CN 107533356 A	02-01-2018
		EP 3268096 A1	17-01-2018
		US 2016267699 A1	15-09-2016
		WO 2016145129 A1	15-09-2016
US 2017351253 A1	07-12-2017	CN 106094861 A	09-11-2016
		US 2017351253 A1	07-12-2017

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82